

## CORONAVIRUS

# Probing the biophysical constraints of SARS-CoV-2 spike N-terminal domain using deep mutational scanning

Wenhao O. Ouyang<sup>1†</sup>, Timothy J.C. Tan<sup>2†</sup>, Ruipeng Lei<sup>1</sup>, Ge Song<sup>3,4,5</sup>, Collin Kieffer<sup>6</sup>, Raiees Andrabi<sup>3,4,5</sup>, Kenneth A. Matreyek<sup>7</sup>, Nicholas C. Wu<sup>1,2,8,9\*</sup>

Increasing the expression level of the SARS-CoV-2 spike (S) protein has been critical for COVID-19 vaccine development. While previous efforts largely focused on engineering the receptor-binding domain (RBD) and the S2 subunit, the amino-terminal domain (NTD) has been long overlooked because of the limited understanding of its biophysical constraints. In this study, the effects of thousands of NTD single mutations on S protein expression were quantified by deep mutational scanning. Our results revealed that in terms of S protein expression, the mutational tolerability of NTD residues was inversely correlated with their proximity to the RBD and S2. We also identified NTD mutations at the interdomain interface that increased S protein expression without altering its antigenicity. Overall, this study not only advances the understanding of the biophysical constraints of the NTD but also provides invaluable insights into S-based immunogen design.

## INTRODUCTION

The emergence of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) has led to the coronavirus disease 2019 (COVID-19) pandemic (1, 2). As the major antigen of SARS-CoV-2, spike (S) glycoprotein plays a critical role in facilitating virus entry (3, 4). Therefore, antibodies to SARS-CoV-2 S are often neutralizing (5, 6). SARS-CoV-2 S protein consists of an N-terminal S1 subunit, which is responsible for engaging the host receptor angiotensin-converting enzyme 2 (ACE2) via the receptor-binding domain (RBD), as well as a C-terminal S2 subunit, which mediates virus-host membrane fusion (4, 7, 8). The S1 subunit also contains an N-terminal domain (NTD) in addition to the RBD (4, 7). While the RBD is generally considered to be immunodominant, the NTD is also a target of neutralizing antibodies (9–11). Structural studies revealed the presence of an antigenic supersite on the NTD that is frequently mutated in SARS-CoV-2 variants of concern (VOCs) (12–17). Amino acid mutations and indels rapidly accumulate within the NTD during the evolution of SARS-CoV-2 in human, at least partly due to the immune selection pressure (18). On the other hand, antibodies to NTD epitopes that are conserved across VOCs have also been identified (16, 19). Despite the importance of NTD in immune response against SARS-CoV-2, the biophysical constraints of NTD remain largely elusive.

COVID-19 vaccines, including both recombinant protein-based and mRNA-based, are proven to be highly protective against SARS-CoV-2 infection (20–23). There is an inverse relationship between the production yield and cost of recombinant protein-based COVID-19

vaccines, such as that from Novavax, which showed promising results in phase 3 clinical trials (22), as well as others that are in earlier phases of clinical trials (24). High protein expression level is also believed to be critical for the effectiveness of mRNA vaccines (25). As a result, identifying mutations that increase S protein expression are crucial for optimizing COVID-19 vaccines. While most studies focused on mutating the S2 subunit and the RBD to increase S protein expression (7, 26–29), little effort has been spent on NTD because of the lack of understanding of its biophysical properties.

Phenotypes of numerous mutations can be measured in a massively parallel manner using deep mutational scanning, which combines saturation mutagenesis and next-generation sequencing (30). Previous studies have applied deep mutational scanning to evaluate the effects of RBD mutations on protein expression, ACE2-binding affinity, and antibody escape (31–36). Although deep mutational scanning of the RBD provided important insights into immunogen design and SARS-CoV-2 evolution (29, 31, 32, 35, 36), similar studies on other regions of the S protein have not yet been carried out.

Here, we used deep mutational scanning to quantify the effects of thousands of NTD single mutations on S protein expression. One notable observation was that NTD residues, unlike RBD residues, showed a weak correlation between mutational tolerability and relative solvent accessibility (RSA). Instead, the mutational tolerability of NTD residues strongly correlated with their distance to RBD and S2. Residues S50 and G232 were two exceptions, in which they were proximal to S2 and RBD, respectively, and yet had a high mutational tolerability. Subsequently, we functionally characterized two mutations that increased S protein expression, namely, S50Q and G232E. These results have important implications toward understanding NTD evolution and S-based immunogen design.

## RESULTS

### Most NTD mutations have minimal impact on S protein expression

To study how SARS-CoV-2 S protein expression is influenced by NTD mutations, we created a mutant library that contained all possible single amino acid mutations across residues 14 to 301 of the S protein. Each of these 288 residues was mutated with the choice of all 19 other

Copyright © 2022  
The Authors, some  
rights reserved;  
exclusive licensee  
American Association  
for the Advancement  
of Science. No claim to  
original U.S. Government  
Works. Distributed  
under a Creative  
Commons Attribution  
License 4.0 (CC BY).

<sup>1</sup>Department of Biochemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. <sup>2</sup>Center for Biophysics and Quantitative Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. <sup>3</sup>Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA 92037, USA. <sup>4</sup>IAVI Neutralizing Antibody Center, The Scripps Research Institute, La Jolla, CA 92037, USA. <sup>5</sup>Consortium for HIV/AIDS Vaccine Development (CHAVD), The Scripps Research Institute, La Jolla, CA 92037, USA. <sup>6</sup>Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. <sup>7</sup>Department of Pathology, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA. <sup>8</sup>Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. <sup>9</sup>Carle Illinois College of Medicine, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.

\*Corresponding author. Email: nicwu@illinois.edu

†These authors contributed equally to this work.

amino acids and the stop codon, leading to a mutant library with 5760 single amino acid mutations. We used a cassette-based polymerase chain reaction (PCR) strategy that was designed to avoid potential off-target errors and generation of double mutants during the library construction process (fig. S1). As a quality control, Sanger sequencing was performed on 10 selected colonies following the construction of the mutant library. All of them contained a single mutation as expected. The mutant library was expressed using the human embryonic kidney (HEK) 293T landing pad cell system such that each transfected cell stably expressed only one mutant (37, 38). Fluorescence-activated cell sorting (FACS) was then performed using the human anti-S2 antibody CC40.8 (39), with phycoerythrin (PE) anti-human immunoglobulin G (IgG) Fc as the secondary antibody. Four separated gates were set up on the basis of the PE signals, each covering 25% of the entire population (fig. S2). The frequency of each mutant among the entire population was calculated (see Materials and Methods), and a cutoff of 0.0075% was set up to filter out mutants with potentially noisy measurements. Among the 5760 missense and nonsense mutations, 3999 (69%) of them satisfied the frequency cutoff for downstream analysis. Notably, the design of our mutant library adopted an internal barcoding strategy that uses synonymous mutations to facilitate sequencing error correction (40). As described previously (41), the expression score of each mutation was calculated on the basis of their frequency in each of the four gates and normalized such that the average expression score of silent mutations was 1 and that of nonsense mutations was 0. Blocks of missing sites could be observed periodically across the NTD, potentially due to the locations of NNK codons in our cassette PCR design (fig. S3).

To evaluate the quality of the deep mutational scanning results, we assessed the expression score distributions of missense, nonsense, and silent mutations (fig. S4A). The difference between the expression scores of silent mutations and nonsense mutations was apparent and significant ( $P = 6 \times 10^{-166}$ ), which validated the selectivity of the deep mutational scanning experiment. Silent mutation and missense mutations had similar expression scores, although the difference is statistically significant ( $P = 2 \times 10^{-5}$ ), indicating that most amino acid mutations in the NTD did not affect S protein expression. While dead cells could potentially increase the noise of the deep mutational scanning results, this effect seemed to be relatively minor in our experiment (fig. S5). In addition, a Pearson correlation of 0.53 was obtained between the expression scores of each mutant from two independent biological replicates (fig. S4B), demonstrating the reproducibility of the deep mutational scanning experiment.

To summarize the expression scores for individual mutations, a heatmap was generated (Fig. 1). We noticed that high-expressing mutations were enriched within the five NTD loop regions (fig. S7A) (12). High-expressing mutations were also found in residues outside of the loop regions, such as residues S50 and G232. This observation shows that some NTD mutations can improve the expression of S protein.

### Mutational tolerability has minimal correlation with solvent accessibility

While some residues were enriched in high-expression mutations (see above), others were enriched in low-expression mutations (e.g., residues D40, L84, and N234; Fig. 1). Consequently, we aimed to identify the biophysical determinants of mutational tolerability in terms of S protein expression. For each residue, we defined the

mutational tolerability as the mean expression score of mutations. A higher mutational tolerability would indicate the enrichment of high-expressing mutations at the specified residue. In contrast, a lower mutational tolerability would indicate the enrichment of low-expressing mutations at the specified residue. A total of 243 NTD residues had six or more mutations with expression score available and were included in this analysis. Notably, a Pearson correlation of 0.68 was obtained between the mutational tolerability values of each position from the two independent biological replicates (fig. S4C).

First, we investigated whether a correlation existed between the mutational tolerability and RSA. Because buried residues are typically important for protein folding stability, residues with a lower RSA are generally expected to have a lower mutational tolerability. For example, previous deep mutational scanning studies on the RBD have shown a decent correlation between RSA and mutational tolerability (Spearman correlation = 0.73; Fig. 2A) (34, 42). In contrast, the mutational tolerability of NTD residues had a much weaker correlation with RSA (Spearman correlation = 0.19; Fig. 2B). These observations indicate that the folding stability of NTD does not have a strong influence on its mutational tolerability and, hence, the S expression level. Alternatively, it is possible that some mutations can destabilize NTD, but NTD instability does not affect the S expression level.

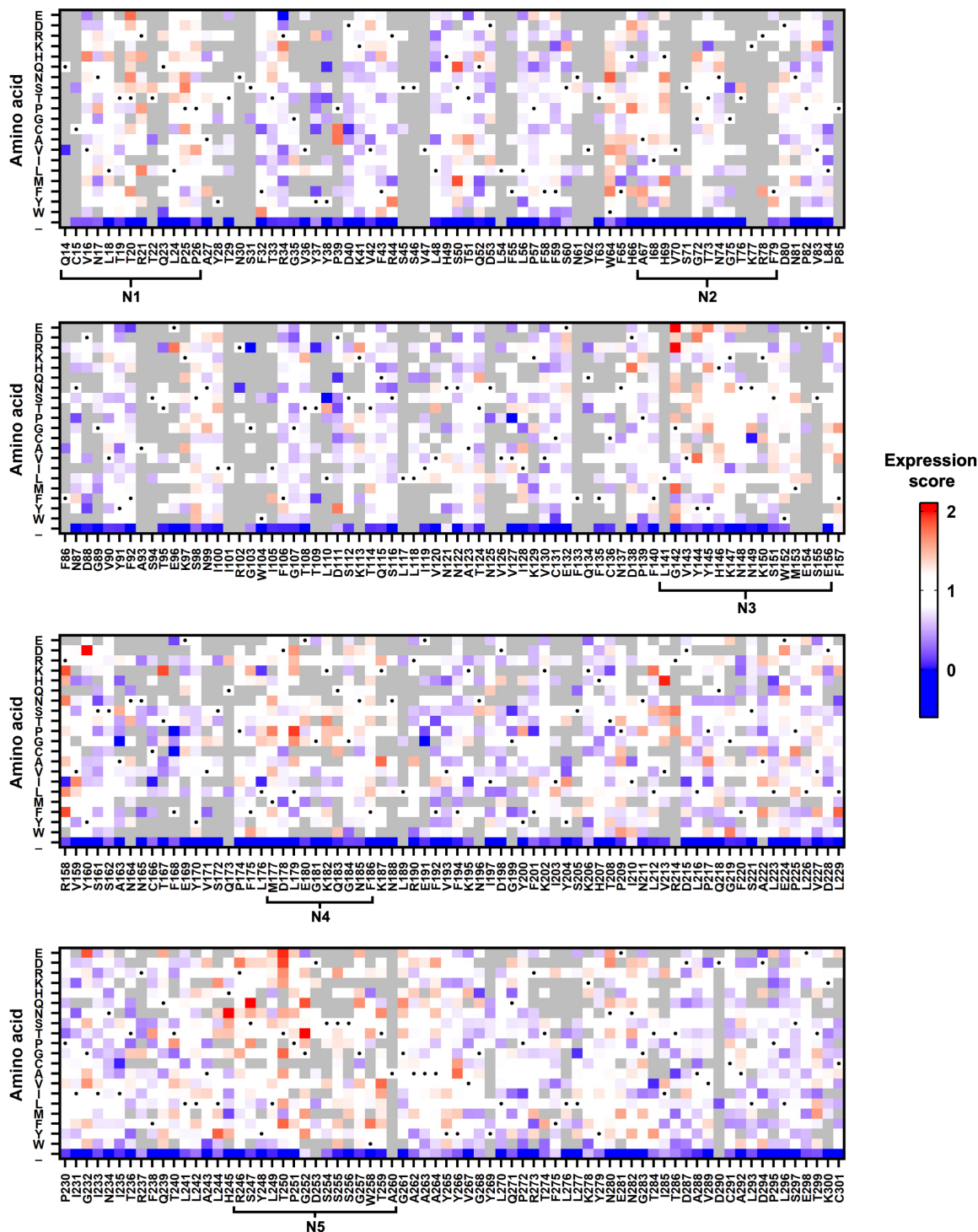
To investigate whether the mutational tolerability correlated with sequence conservation, we then analyzed the NTD sequences of 27 sarbecovirus strains, including SARS-CoV-2. Less conserved residues tended to have a higher mutational tolerability, while more conserved residues tended to have a lower mutational tolerability, although the correlation was not strong (Spearman correlation = -0.30; Fig. 2C). In comparison, the correlation between sequence conservation and RSA was even weaker (Spearman correlation = -0.16; Fig. 2D).

### Mutational tolerability correlates with distance to RBD/S2

We further calculated the distance from each NTD residue to RBD/S2 of the S protein. A positive correlation was observed between the mutational tolerability and the distance to RBD/S2 (Spearman correlation = 0.55; Fig. 2E). In other words, the more distant an NTD residue was from the RBD/S2, the higher the mutational tolerability was. This correlation was apparent when the mutational tolerability of each NTD residue was projected on the S protein structure (Fig. 2G). Consistently, the epitopes of two cross-neutralizing antibodies, namely, C1717 and C1791, were significantly closer to RBD/S2 ( $P \leq 1 \times 10^{-4}$ ) and had lower mutational tolerability ( $P \leq 0.03$ ) when compared to the rapidly evolving NTD antigenic supersites (Fig. 2F and fig. S7B) (14, 16, 43).

### Naturally circulating NTD indel sites have significantly higher mutational tolerability

We then examined the naturally occurring NTD mutations and indels observed among 17 SARS-CoV-2 major variants using our deep mutational scanning data. Among these 17 variants, there are 25 different amino acid mutations and 25 indel sites relative to the ancestral strain. Among the 25 amino acid mutations and 25 indel sites, 20 and 23, respectively, have available expression scores and site-wise mutational tolerability in our dataset. The expression score distribution of the 20 natural amino acid mutations was similar to the rest of the missense mutations ( $P = 0.15$ ; Fig. 3A). In contrast, the 23 natural indel sites had significant higher mutational

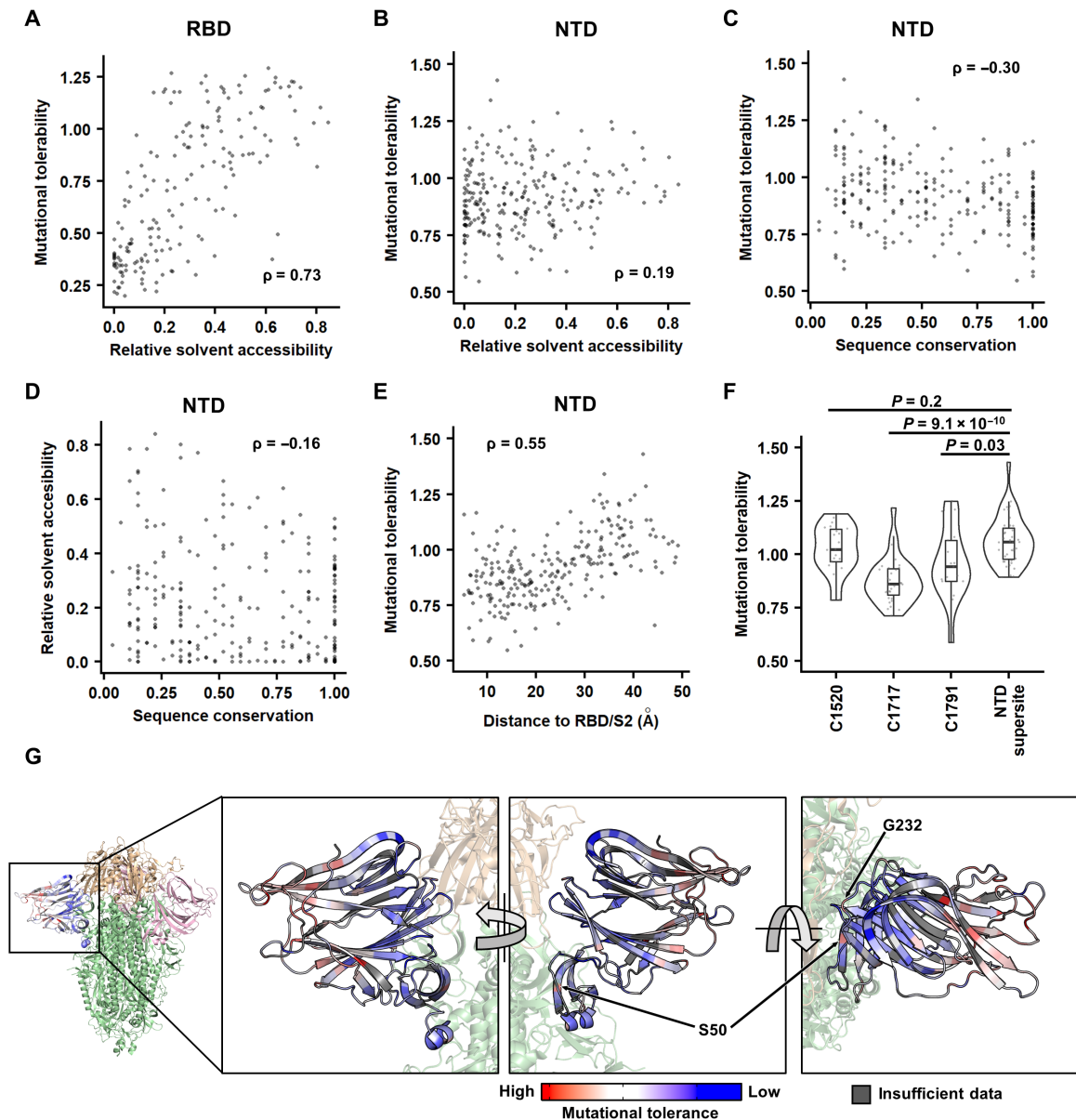


**Fig. 1. Effects of NTD single mutations on S protein expression.** The expression scores of individual NTD mutations are shown as a heatmap. X axis represents the residue position. Y axis represents different amino acids and the stop codon (⊔). Amino acids corresponding to the wild-type (WT) sequence are indicated by the black dots. Mutations with a total frequency of <0.0075% were excluded from the analysis and shown in gray. Regions corresponding to the N1-N5 loops were defined as previously described (12).

tolerability than the other NTD residues ( $P = 2 \times 10^{-7}$ ; Fig. 3B). This observation is consistent with the enrichment of natural indels in the five NTD loops, which have high mutational tolerability (Fig. 3, C and D).

**Two buried NTD mutations increase S protein expression**

While NTD residues adjacent to RBD/S2 typically had a low mutational tolerability, S50 and G232 were two exceptions (Fig. 2G). For example, mutations S50G and G232E had a high expression score in



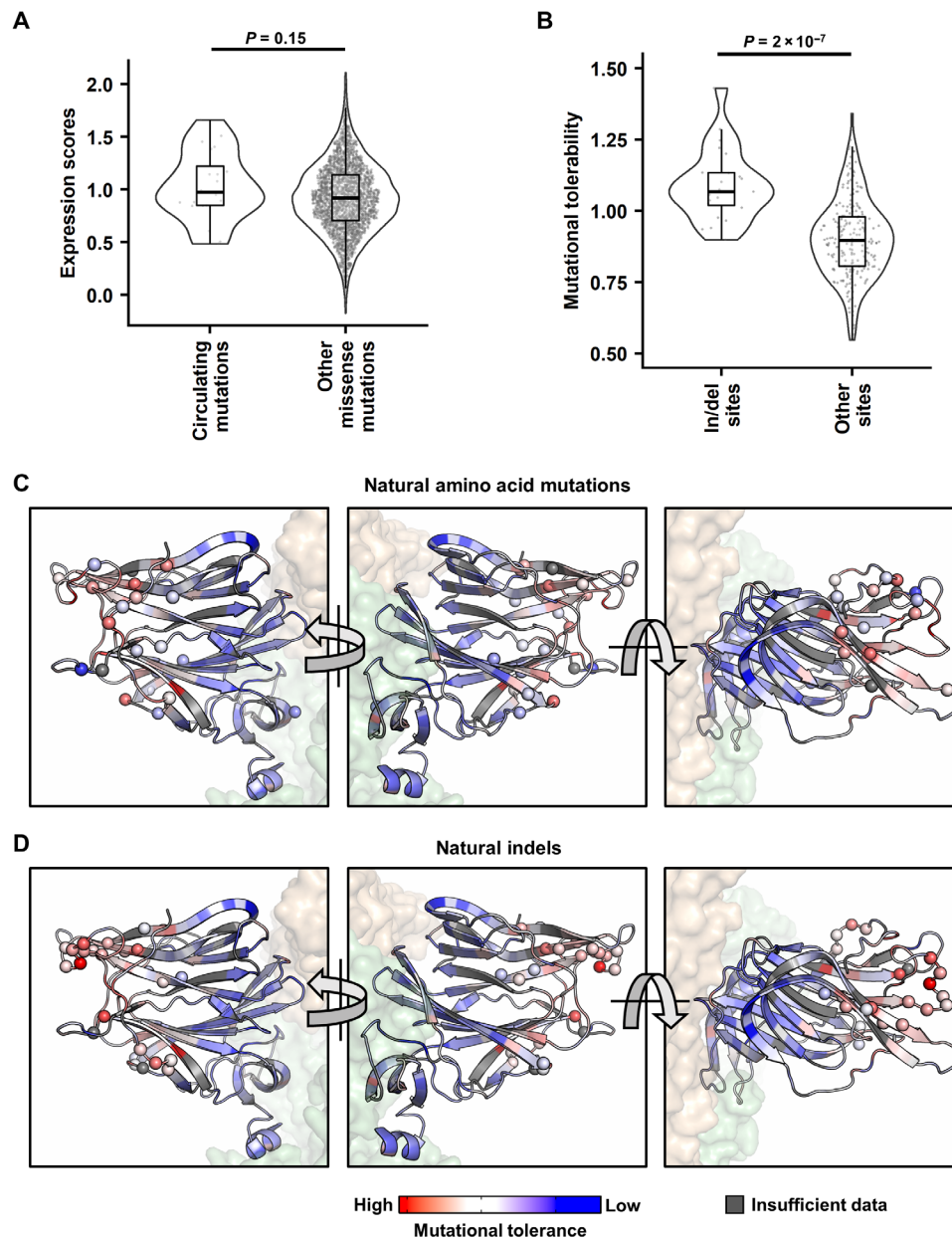
**Fig. 2. Identifying the biophysical determinants of mutational tolerability.** (A and B) The relationship between RSA and the mutational tolerability is shown for (A) RBD and (B) NTD. The deep mutational scanning data on RBD expression were from a previous study (42). (C and D) Relationship between sequence conservation among 27 sarbecovirus strains (table S6) and (C) the mutational tolerability or (D) RSA of each NTD residue. (E) Relationship between the distance to RBD/S2 and the mutational tolerability of each NTD residue. (A to E) Each data point represents one residue. The Spearman's rank correlation coefficient ( $\rho$ ) is indicated. (F) The mutational tolerability of residues within the cross-neutralizing NTD antibody epitopes (C1520, C1717, and C1791) (16) is compared to that within the antigenic supersite (14) using a violin plot. Each data point represents one residue.  $P$  values were computed by two-tailed  $t$  test. (G) The mutational tolerability of each NTD residue is projected on one NTD of the S trimer structure [Protein Data Bank (PDB) 6ZGE (45) and PDB 7B62 (46)]. Red indicates residues with higher mutational tolerability, while blue indicates residues with lower mutational tolerability. Residues with insufficient data to calculate mutational tolerability are colored in gray. Two residues of interests, namely, S50 and G232, are indicated. RBDs are colored in wheat, the two other NTDs are in pink, and the rest of the S1 and S2 subunits are in green.

our deep mutational scanning results. To validate this finding, we used the same landing pad system to construct HEK293T cell lines that stably expressed S50Q, G232E, and S50Q/G232E double mutant. As quantified by flow cytometry analysis (Fig. 5A and fig. S8), the mean fluorescence intensity (MFI) of S50Q and G232E increased from wild type (WT) by 1.7-fold ( $P = 0.002$ ) and 1.5-fold ( $P = 9 \times 10^{-4}$ ), respectively, whereas that of S50Q/G232E increased by 2.5-fold ( $P = 2 \times 10^{-6}$ ). Although fold change in MFI is unlikely to linearly

correlate with the fold change in protein synthesis, such increase in MFI shows that S50Q, G232E, and S50Q/G232E double mutant have increased surface expression compared to WT.

Subsequently, we examined the natural occurrences of S50Q and G232E mutations. Both S50Q and S232E rarely occur in circulating SARS-CoV-2. Among over 10 million NTD sequences on Global Initiative for Sharing Avian Influenza Data (GISAID) (44), only 22 and 3 sequences contain S50Q and G232E, respectively. To probe





**Fig. 3. Analysis of naturally circulating NTD mutations.** (A) Expression scores of naturally occurring NTD mutations and all other missense mutations. (B) Mutational tolerability of naturally occurring indel sites and all other sites.  $P$  values were computed by two-tailed  $t$  test. (C and D) Missense mutations (C) and indel sites (D) observed in 17 SARS-CoV-2 major variants are labeled as spheres on one NTD of the S trimer structure [PDB 6ZGE (45) and PDB 7B62 (46)]. Red indicates residues with higher mutational tolerability, while blue indicates residues with lower mutational tolerability. Residues with insufficient data to calculate mutational tolerability are colored in gray. The neighboring RBDs are colored in wheat, and the rest of the S1 and S2 subunits shown are colored in green.

the structural impact of S50Q and G232E, we analyzed their local environments on the structure of S protein and performed structural modeling using Rosetta (Fig. 5, C and D) (45–47). S50 forms a hydrogen bond with K304 and is proximal to the S2 subunit. Structural modeling showed that S50Q not only is able to maintain the hydrogen bond with K304 but also strengthens the van der Waals interaction between the NTD and S2 by pushing K304 toward S2 from the adjacent protomer (Fig. 5C). G232 is proximal to a positively charged region on the RBD that is featured by R355 and R466 (Fig. 5D). Structural modeling suggested that G232E could form

favorable electrostatic interactions with both R355 and R466. We further recombinantly expressed these mutants and tested their thermostability using a thermal shift assay (Fig. 5B). Notably, all the recombinantly expressed S proteins contained K986P/V987P mutations in the S2 subunit, which are known to stabilize the prefusion conformation and increase expression (26, 48). The melting temperatures of WT and NTD mutants were almost identical at a  $T_m$  of 46°C to 46.5°C. These observations indicate that despite both S50Q and G232E improve the interaction between NTD and the rest of the S protein, they have minimal impact on the global folding stability

of the S protein. Notably, all mutants increased the yield of the soluble recombinantly expressed S protein compared to WT (fig. S6), although the rank order was different from that of the membrane-bound form (Fig. 4A and fig. S6).

### S50Q and G232E have minimal effects on the fusion activity and antigenicity

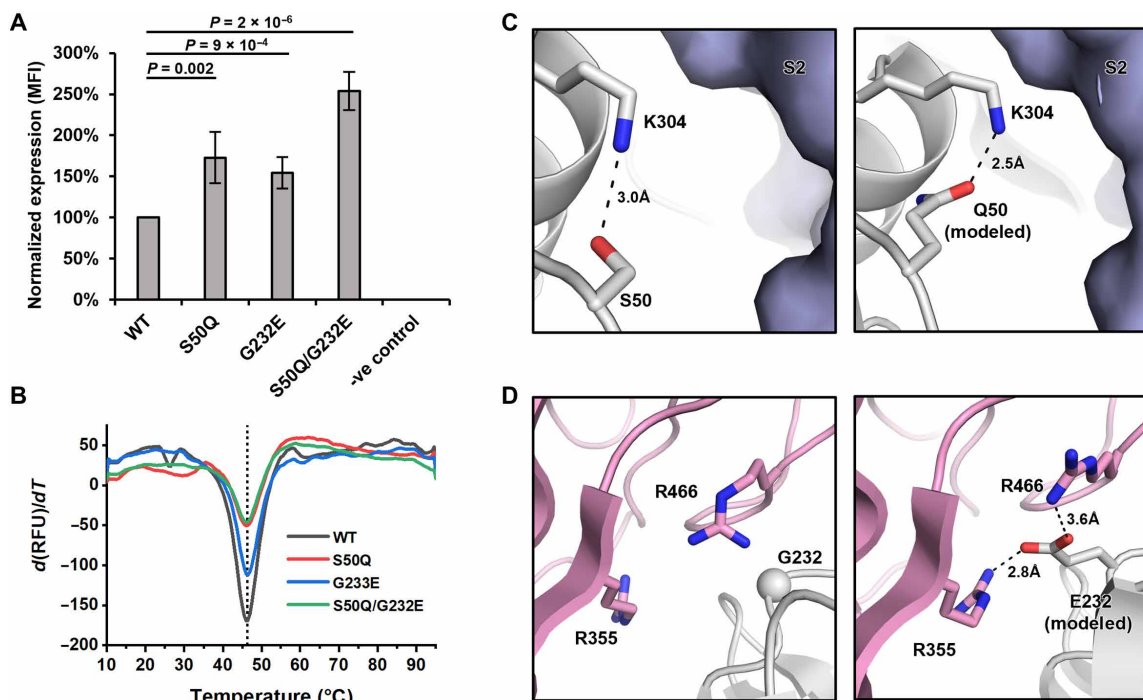
To understand the functional consequences of S50Q and G232E, we further tested whether S50Q, G232E, and S50Q/G232E exhibited a change in fusion activity compared to WT. A fluorescence-based cell-cell fusion assay that relied on the split mNeonGreen2 (mNG2) (49) was performed (see Materials and Methods; fig. S9). Briefly, HEK293T landing pad cells that expressed human ACE2 (hACE2) and mNG2<sub>1-10</sub> were mixed with HEK293T landing pad cells that expressed S proteins and mNG2<sub>11</sub>. Green fluorescence due to mNG2 complementation was generated when fusion between the two cell lines occurred. Fluorescence microscopy analysis showed that all mutants facilitated hACE2-mediated fusion (Fig. 6, A and B). Consistently, flow cytometry analysis at both 3-hour and 24-hour postmixing indicated that none of the tested mutants diminished the fusion activity when compared to WT (Fig. 6, C and D). At 3-hour postmixing, both S50Q (24%,  $P = 0.03$ ) and G232E (25%,  $P = 0.01$ ) showed mild, yet significant, increases in fusion activity compared to WT. Similarly, at 24-hour postmixing, S50Q (19%,  $P = 0.01$ ), G232E (13%,  $P = 0.02$ ), and S50Q/G232E double mutant (37%,  $P = 0.005$ ) all showed an increase in fusion activity compared to WT. Such a mild increase in fusion activity may simply be attributed

to the higher expression level of the mutants. However, because the relationship between protein expression level and fusion activity remained elusive, we are unable to assess the fusion activity per S protein molecule based on the current data. Negative control cells expressing the K986P/V987P double mutant, which is known to stabilize the prefusion form of the S protein (26, 48), did not show any fusion activity (Fig. 5, A to D). Overall, our results demonstrated that both S50Q and G232E did not affect the fusogenic capabilities of the S protein.

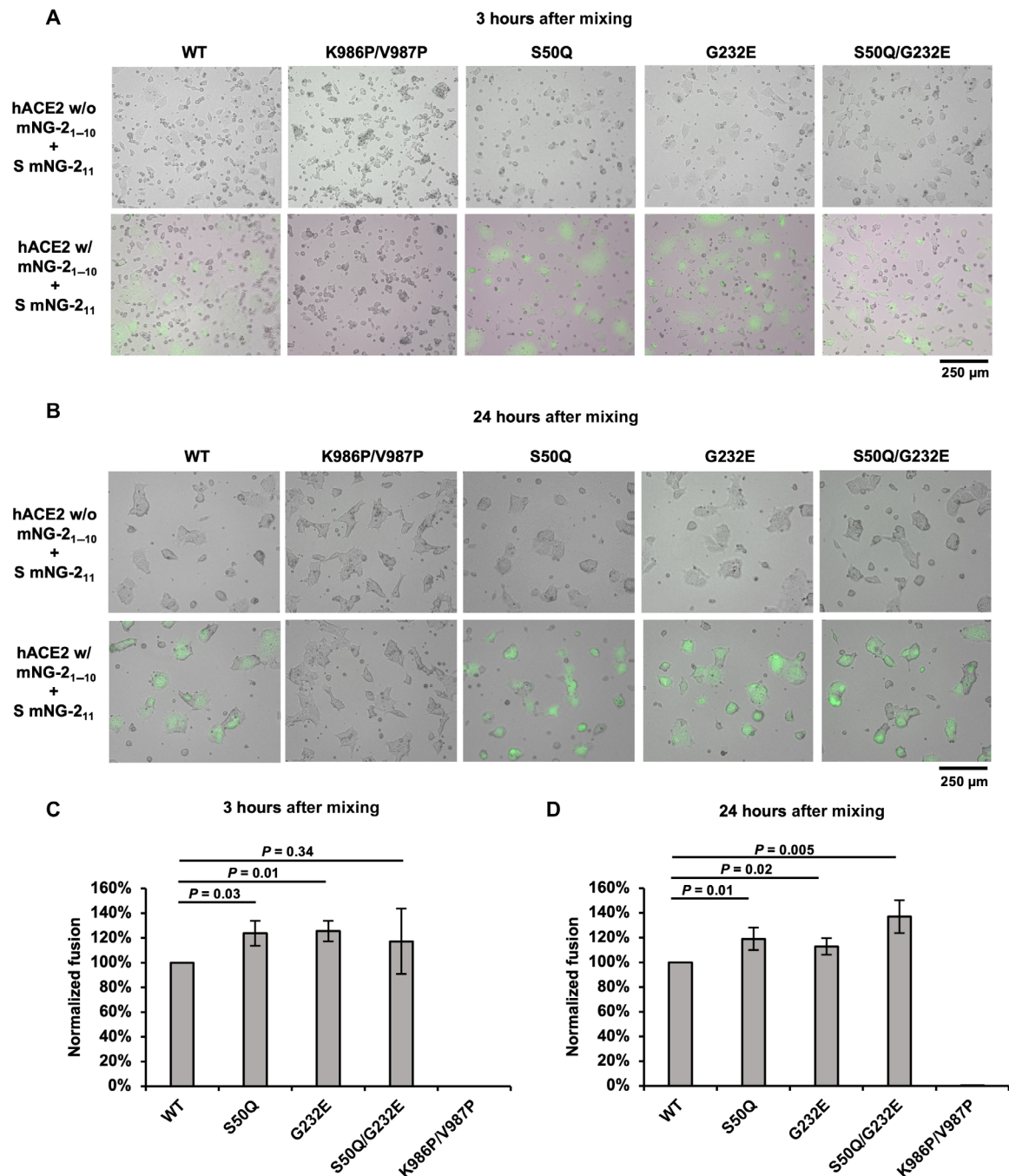
We then proceeded to investigate whether S50Q, G232E, and S50Q/G232E alter the antigenicity of the S protein. The binding of three antibodies targeting different domains of the S protein was tested, namely, CC12.3 (anti-RBD) (50), S2M28 (anti-NTD) (14), and COVA1-07 (anti-S2) (51). Notably, S2M28 is an NTD supersite-targeting antibody. Flow cytometry analysis showed that all three antibodies bound to the tested mutants at a similar level to WT (Fig. 6 and fig. S10), indicating that S50Q, G232E, and S50Q/G232E did not alter the structural conformation and antigenicity of the S protein.

### DISCUSSION

S protein is central to the research of SARS-CoV-2 evolution and COVID-19 vaccines (50, 52–54). While both the RBD and the NTD on the S protein are targets of neutralizing antibodies and are involved in the antigenic drift of SARS-CoV-2 (43, 55–61), the NTD often receives less attention than does the RBD. Using deep mutational scanning, this study shows that many NTD mutations at buried



**Fig. 4. S50Q and G232E at the interdomain interface increase S protein expression.** (A) Cell surface S protein expression of WT and NTD mutants was quantified using flow cytometry analysis with CC40.8 as the primary antibody. Untransfected HEK293T landing pad cells were used as a negative control (-ve control). S protein expression level was defined as the MFI of the positive gated population. S protein expression level was normalized to WT. The error bar indicates the SD of six independent biological replicates.  $P$  values were computed by two-tailed  $t$  test. (B) Thermostability of WT S protein and selected NTD mutants was measured using differential scanning fluorimetry. The black vertical dotted line indicates the melting temperature of WT ( $T_m = 46.2^\circ\text{C}$ ). (C and D) Rosetta-based structural modeling of (C) S50Q and (D) G232E was performed using the structure of S protein (PDB 6ZGE) (46). The three protomers of the S protein are colored in white, light blue, and pink. Potential interactions are represented by black dashed lines with distance labeled.



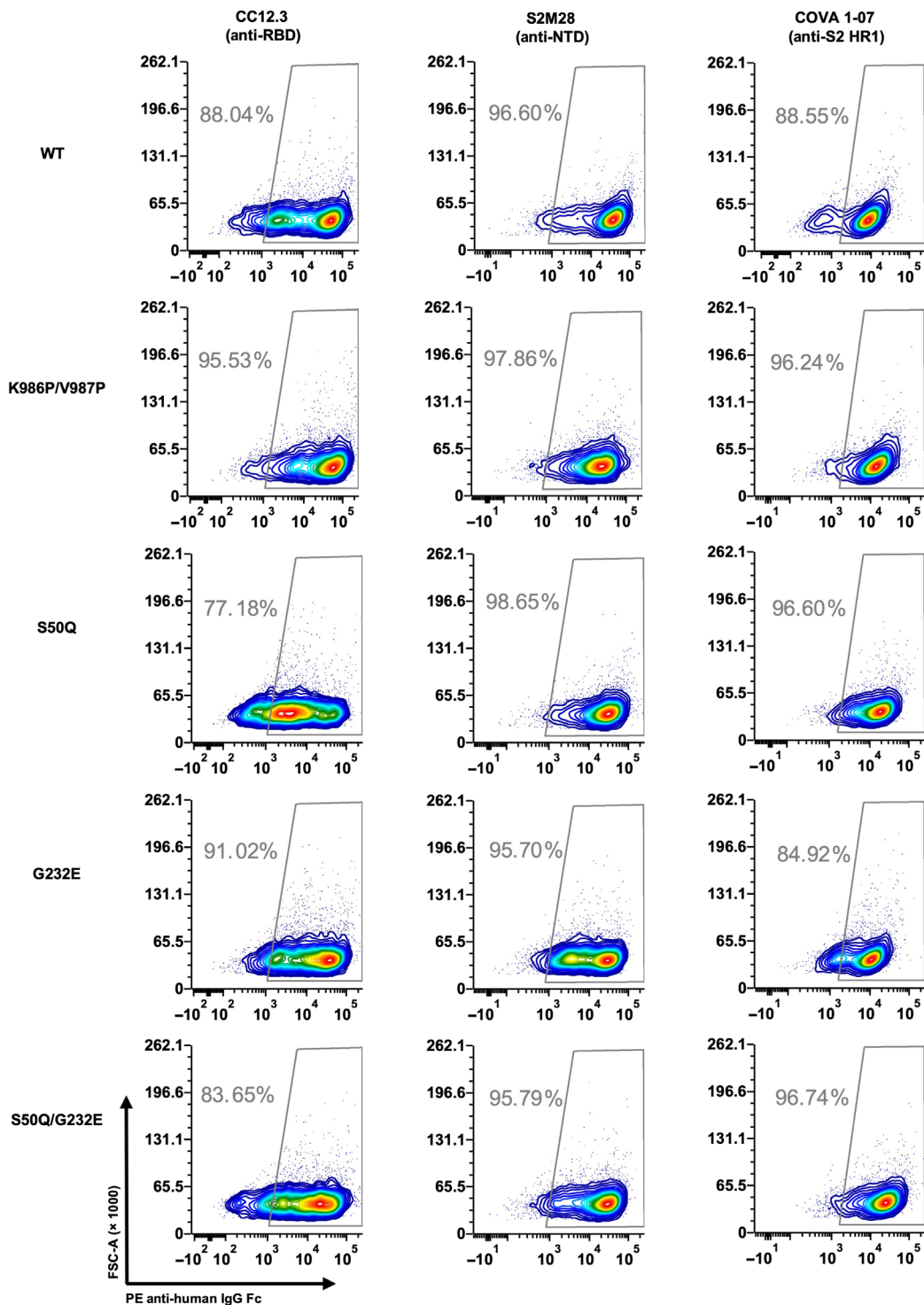
**Fig. 5. S50Q, G232E, and S50Q/G232E do not diminish fusion activity.** (A and B) Fluorescence microscopy analysis of the fusion events at (A) 3-hour and (B) 24-hour postmixing of cells expressing S and mNG2<sub>11</sub> (S cells) and cells expressing hACE2 and mNG2<sub>1-10</sub> (hACE2 cells). Cells with green fluorescence signals are the fused cells. Scale bars are shown at the bottom right corner. (C and D) Fusion activity of WT and selected NTD mutants at (C) 3-hour and (D) 24-hour postmixing was quantified using flow cytometry analysis. Fusion activity was normalized to WT. The error bar indicates the SD of at least four biological replicates. *P* values were computed by two-tailed *t* test.

residues do not affect S protein expression. At the same time, the closer an NTD mutation is to RBD/S2, the more likely it is detrimental to S protein expression. These observations imply that for optimum S protein expression, the structural stability at the NTD-RBD and the NTD-S2 interfaces is more critical than the folding stability of the NTD. Our results also at least partly explain why the N1 to N5 loops, which contain the NTD antigenic supersite (62) and

are far from the NTD-RBD/S2 interfaces, are highly diverse among SARS-CoV-2 variants and sarbecovirus strains. Overall, this study provides crucial biophysical insights into the evolution of the NTD.

NTD mutations S50Q and G232E, which locate at the inter-domain interface and increase S protein expression, represent another important finding of this study. Engineering high-expressing S protein can lower the production cost of recombinant COVID-19 vaccine





**Fig. 6. S50Q, G232E, and S50Q/G232E do not alter antibody binding.** Three antibodies targeting different domains on S were tested for binding to cells expressing WT, K986P/V987P, S50Q, G232E, or S50Q/G232E S protein. Binding was measured by flow cytometry analysis. Gating was set up using untransfected HEK293T landing pad cells, which served as a negative control (fig. S10).

and may improve the effectiveness of mRNA vaccines (25). Similar to certain previously characterized mutations in the S2 (26, 27), S50Q and G232E in the NTD increase the expression yield of the S protein without changing its  $T_m$ . Consistently, a recent study showed that

NTD mutations in BA.1 improve the expression of S protein without increasing its thermostability (63). Furthermore, S50Q and G232E are not solvent-exposed on the S protein surface and do not seem to alter the antigenicity of the S protein. Notably, according to our



deep mutational scanning data, S50Q and G232E are just two of many mutations that enhance S protein expression. Therefore, although most studies on S-based immunogen design focus on the mutations in the RBD and S2 (7, 26–29), our results suggest that mutations in NTD can provide a complementary strategy.

We acknowledge that the S protein expression level does not necessarily correlate with virus replication fitness. For example, NTD mutations that do not affect the S protein expression may be detrimental to the replication fitness of SARS-CoV-2 due to the negative impact on NTD functionality. While the functional importance of the NTD in natural infection remains largely unclear, NTD has been proposed to facilitate virus entry by interacting with DC-SIGN, L/SIGN, AXL, ASGR1, and KREMEN1 (64–66). Studies have also shown that the NTD can allosterically evade antibody binding by interacting with a heme metabolite (46) and modulate the efficiency of virus-host membrane fusion (67, 68). To fully comprehend the biophysical constraints of NTD, future studies should systematically investigate how different NTD mutations affect virus replication fitness.

## MATERIALS AND METHODS

### Construction of the NTD mutant library

SARS-CoV-2 S NTD mutant library was constructed based on the HEK293T landing pad system (37, 38). The template for constructing the NTD mutant library was a plasmid that encoded (from 5' to 3') an attB site, a codon-optimized SARS-CoV-2 S (GenBank ID: NC\_045512.2) with the PRRA motif in the furin cleavage site deleted, an internal ribosome entry site (IRES), and a puromycin resistance marker. This plasmid was used as a PCR template to generate a linearized vector and a library of mutant NTD inserts. The linearized vector was generated using 5'-TGCTCGTCTCTACAACCTCCGC-CAGCTTCAGCACC-3' and 5'-TGCTCGTCTCTTCACTGGC-CGTCGTTTTACAACG-3' as primers. Inserts were generated by two separate batches of PCRs to cover the entire NTD. The first batch of PCRs consisted of 36 reactions, each containing one cassette of forward primers and the universal reverse primer 5'-TGCTCGTCTC-GTTGTACAGCACGGAGTAGTCGGC-3'. Each cassette contained an equal molar ratio of eight forward primers that had the same 21 nucleotides (nt) at the 5' end and 15 nt at the 3' end. Each primer within a cassette was also encoded with an NNK (N: A, C, G, T; K: G, T) sequence at a specified codon positions for saturation mutagenesis. In addition, each primer also carried unique silent mutations (also known as synonymous mutations) to help distinguish between sequencing errors and true mutations in downstream sequencing data analysis as described previously (40). The forward primers, named as CassetteX\_N (X: cassette number, N: primer number), are listed in table S1. The second batch of PCR consisted of another 36 PCRs, each with a universal forward primer 5'-TGCTCGTCTCAGTGAATTGTAATACGACTCACTA-3' and a unique reverse primer as listed in table S2. Subsequently, 36 overlapping PCRs were performed using the universal forward and reverse primers, as well as a mixture of 10 ng each of the corresponding products from the first and second batches of PCR. The 36 overlap PCR products were then mixed at equal molar ratio to generate the final insert of the NTD mutant library. All PCRs were performed using PrimeSTAR Max polymerase (Takara Bio, catalog no. R045B) per the manufacturer's instruction, followed by purification using the Monarch Gel Extraction Kit (New England Biolabs, catalog no. T1020L). The final insert and the linearized vector were digested by BsmBI-v2

(New England Biolabs, catalog no. R0739L) and ligated using T4 DNA Ligase (New England Biolabs, catalog no. M0202L). Ligation product was purified by the PureLink PCR Purification Kit (Thermo Fisher Scientific, catalog no. K310002) and then transformed into MegaX Dh10B T1R cells (Thermo Fisher Scientific, catalog no. C640003). At least half a million colonies were collected. Plasmid mutant library was purified from the bacteria colonies using the PureLink HiPure Plasmid Midiprep Kit (Invitrogen, catalog no. K210005). All primers in this study were ordered from Integrated DNA Technologies.

### Construction of stable cell lines using HEK293T landing pad cells

HEK293T landing pad cells (37, 38) were used to display the NTD mutant library for deep mutational scanning. Landing pad cells were maintained using complete growth medium consisting of Dulbecco's modified Eagle's medium (DMEM; Corning), 10% (v/v) fetal bovine serum (FBS; VWR), penicillin-streptomycin (Gibco), nonessential amino acid (Gibco), and doxycycline (2 µg/ml). Plasmid (1.2 µg) was transfected into  $6 \times 10^5$  landing pad cells. For the deep mutational scanning experiment, eight transfection reactions were carried out in parallel to minimize loss of mutant diversity at the transfection step. Transfected cells were then incubated at 37°C with 5% CO<sub>2</sub>. After 48 hours, 10 nM AP1903 was supplemented to carry out negative selection. At 72 hours after the negative selection, positive selection antibiotic [puromycin (1 µg/ml) for NTD cell lines or hygromycin (100 µg/ml) for hACE2 cell lines] was supplemented to the medium to carry out positive enrichment of cells with successful recombination. Constructed cell lines would remain in the complete growth medium supplemented with doxycycline and the positive selection antibiotics.

### Sorting the NTD mutant library based on S protein expression level

Four T-75 flasks (Corning) that were 90% confluent with cells that carried the NTD mutant library were washed with 1× phosphate-buffered saline (PBS), harvested with warm versene, and pelleted via centrifugation at 300g for 5 min at room temperature. Cells were then resuspended in FACS buffer [2% (v/v) FBS, 5 mM EDTA in DMEM supplemented with glucose, L-glutamine, and HEPES but without phenol red (Gibco)]. Subsequently, cells were incubated with CC40.8 (5 µg/ml) at 4°C with gentle shaking for 1 hour. Cells were washed once with ice-cold FACS buffer and incubated with PE anti-human IgG Fc (1 µg/ml; BioLegend, catalog no. 410708) at 4°C with gentle shaking in the dark for 1 hour. Cells were washed once and resuspended in ice-cold FACS buffer. Cells were then filtered using a 40-µm cell strainer (VWR) before cell sorting. FACS was performed using a BD FACSAria II cell sorter (BD Biosciences) with a 561-nm laser and a 582/15 band-pass filter. Cells were collected into ice-cold D10 medium [DMEM with glucose (4.5 g/liter), 4 mM L-glutamine, and sodium pyruvate (110 mg/liter; Corning), supplemented with 10% (v/v) FBS (VWR), 1× penicillin-streptomycin (Gibco), and 1× nonessential amino acids (Gibco)] and binned into no (bin 0), low (bin 1), medium (bin 2), and high (bin 3) expression according to PE signal, where each bin contains 25% of the singlet population (fig. S2). A biological replicate of the deep mutational scanning experiment was performed, starting from the transfection step.

### Next-generation sequencing of the NTD mutant library

Sorted cells from each bin were pelleted at 300g, 4°C for 15 min and then resuspended in 200 µl of PBS (Corning). Genomic DNA

extraction was then performed using the DNA Blood and Tissue Kit (Qiagen, catalog no. 69504) according to the manufacturer's instructions with a modification: Cells were incubated at 56°C for 30 min instead of 10 min. The NTD mutant library was amplified from the genomic DNA in two nonoverlapping fragments using KOD Hot Start DNA polymerase (MilliporeSigma, catalog no. 710863) per the manufacturer's instruction with the following two primer sets, respectively (also see table S3): set 1: 5'-CACTCTTCCCTACACGACGCTCTTCCGATCTCTGCTGCCTCTGGTGTCCAGC-3' (NTD-DMS-recover-1F) and 5'-GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTTGGCGCTGCTGTACACCCG-3' (NTD-DMS-recover-1R); set 2: 5'-CACTCTTCCCTACACGACGCTCTTCCGATCTAGCTGGATGGAAAGCGAGTTC-3' (NTD-DMS-recover-2F) and 5'-GACTGGAGTTCAGACGTGTGCTCTTCCGATCTCACGGTGAAGGACTTCAGGGT-3' (NTD-DMS-recover-2R).

A second round of PCR was carried out to add the adapter sequence and index to the amplicons as described previously (69). The final PCR products were submitted for next-generation sequencing using Illumina MiSeq PE300.

### Analysis of next-generation sequencing data

Next-generation sequencing data were obtained in FASTQ format. Forward and reverse reads of each paired-end read were merged by PEAR (70). The merged reads were parsed by SeqIO module in BioPython (71). Primer sequences were trimmed from the merged reads. Trimmed reads with lengths inconsistent with the expected length were discarded. The trimmed reads were then translated to amino acid sequences, with sequencing error correction performed at the same time as previously described (40). Amino acid mutations were called by comparing the translated reads to the WT amino acid sequence. Frequency ( $F$ ) of a mutant  $i$  at position  $s$  within bin  $n$  of replicate  $k$  was computed for each replicate as follows

$$F_{i,s,n,k} = \frac{\text{readcount}_{i,s,n,k} + 1}{\sum_s \sum_i (\text{readcount}_{i,s,n,k} + 1)} \quad (1)$$

A pseudocount of 1 was added to the read counts of each mutant to avoid division by zero in subsequent steps. We then calculated the total frequency ( $F_{\text{total}}$ ) of mutant  $i$  at position  $s$  as follows

$$F_{\text{total},i,s} = \frac{\sum_{k=1}^2 \sum_{n=0}^3 F_{i,s,n,k}}{8} \quad (2)$$

Mutants with  $F_{\text{total}}$  of equal or greater than 0.0075% were selected for downstream analysis. Subsequently, the weighted average ( $W$ ) of each mutant among 4 bins (bin 0 to bin 3) in each replicate was computed as described previously (41)

$$W_{i,s,k} = \frac{F_{i,s,0,k} \times 0.25 + F_{i,s,1,k} \times 0.5 + F_{i,s,2,k} \times 0.75 + F_{i,s,3,k} \times 1}{\sum_{n=0}^3 F_{i,s,n,k}} \quad (3)$$

Selected mutants were then categorized based on the mutation types (missense, nonsense, and silent). The mean value of weighted average for nonsense and silent mutations was calculated. Expression score (ES) of a mutant  $i$  at position  $s$  of replicate  $k$  was calculated as described previously (41)

$$ES_{i,s,k} = \frac{W_{i,s,k} - \overline{W_{\text{nonsense},k}}}{\overline{W_{\text{silent},k}} - \overline{W_{\text{nonsense},k}}} \quad (4)$$

The final expression score of a mutant  $i$  at position  $s$  was calculated by taking the average of the expression scores between replicates. Mutational tolerability of position  $s$  was then calculated by taking the average of the expression scores of all mutants at that position

$$\text{mutational tolerability}_s = \frac{\sum_{i \in s} ES_{i,s}}{|\{i \in s\}|} \quad (5)$$

### Structural analysis of deep mutational scanning results

DSSP (72, 73) was used to calculate the solvent exposure surface area (SASA) of each residue in NTD and RBD on the S trimer [Protein Data Bank (PDB) 6ZGE] (45). Deep mutational scanning result of RBD was extracted from a previous study (42). RSA was computed by dividing the SASA by the theoretical maximum allowed solvent accessibility of the corresponding amino acid (74).

Each NTD residue's distance to RBD/S2 was calculated on the basis of the S trimer structure (PDB 6ZGE) (45), with the NTD replaced by the high-resolution crystal structure (PDB 7B62) (46). For each NTD residue, the distances to all RBD and S2 residues were measured. The shortest distance was then recorded as the "distance to RBD/S2." Residue-residue distance was defined as the distance between the centroid coordinates of two residues.

To visualize the mutational tolerability of each NTD residue, the crystal structure of SARS-CoV-2 S protein NTD (PDB 7B62) was used (46). The NTD crystal structure was then aligned with the S trimer to generate the figures (PDB 6ZGE) (45).

### Evolution analysis of NTD sequences

The sequence conservation analysis of NTD was based on 27 sarbecovirus strains (table S6) (1, 75–79). S sequences of these strains were retrieved from GenBank and GISAID (44). Their NTD sequences were then identified using tBlastn search using the amino acid sequence of SARS-CoV-2 Hu-1 NTD (gene ID: 43740568) as the query sequence. The BlastXML output of tBlastn was then parsed and used as the input for multiple sequence alignment using MAFFT (80, 81). For each residue position, sequence conservation was defined as the proportion of strains that contains the same amino acid variant as SARS-CoV-2 Hu-1. The information of the 17 SARS-CoV-2 major variants and the observed naturally occurring NTD missense/indel sites were collected from ViralZone (table S7) (82).

### Rosetta-based mutagenesis

The structure of the S protein was obtained from PDB (PDB 6ZGE) (45). Water molecules and *N*-acetyl-D-glucosamine were removed using PyMOL (Schrödinger). Then, the amino acids were renumbered using pdb-tools (83). Fixed backbone point mutagenesis for S50Q and G232E was performed using the "fixbb" application in Rosetta (RosettaCommons). One-hundred poses were generated for each mutagenesis. Using the lowest-scoring structure from fixed backbone mutagenesis as input, a constraint file was obtained using the minimize\_with\_cst application in Rosetta. Fast relax was then performed via the "relax" application in Rosetta (47) with the corresponding constraint file. The lowest-scoring structure out of eight was then used for structural analysis. Code and source files for structural modeling are available in [https://github.com/nicwulab/SARS-CoV-2\\_NTD\\_DMS/tree/main/rosetta](https://github.com/nicwulab/SARS-CoV-2_NTD_DMS/tree/main/rosetta).

### Split mNG2-based cell-cell fusion assay

hACE2 construct was constructed in a previous study (38). A split mNG2 reporter system was integrated into the S plasmid (see above) and the hACE2 plasmid (49). Specifically, a gene fragment that encoded (from 5' to 3') a GCN4 leucine zipper, a GS linker, mNG2<sub>1-10</sub>, and a 2A self-cleaving peptide was inserted into the hACE2 plasmid between the IRES and the hygromycin resistance marker. Similarly, a gene fragment that encodes (from 5' to 3') a GCN4 leucine zipper, a GS linker, mNG2<sub>11</sub>, and a 2A self-cleaving peptide was inserted into the S plasmid between the IRES and the puromycin resistance marker. Each plasmid construct was transfected and recombined into HEK293T landing pad cells per step described above.

Once the stable cell lines were created,  $5 \times 10^5$  landing pad cells expressing hACE2 with mNG2<sub>1-10</sub> were seeded in six-well plates (Fisher Scientific). The cells were then incubated at 37°C with 5% CO<sub>2</sub> for 15 min to allow seeding. Subsequently,  $5 \times 10^5$  landing pad cells expressing S with mNG2<sub>11</sub> were then added dropwise to the seeded hACE2 cells. Both cells were filtered through 40- $\mu$ m cell strainer (VWR) before seeding. At 3- and 24-hour postmixing, fusion events in each well were qualitatively assessed with an ECHO Revolve epifluorescence microscope (ECHO) in inverted format. Overlaid images were captured on white light and fluorescein isothiocyanate filter channels using an UPlanFL N 10 $\times$ /0.30 numerical aperture objective (Olympus) with identical light intensity and exposure settings for all conditions. Cells in each well were then collected using 0.5 mM EDTA, pelleted via centrifugation at 300g for 5 min at room temperature, and resuspended in the FACS buffer. LSRII flow cytometry (BD Biosciences) was used to quantify the fusion events of each sample. Negative controls were measured first to set up proper gating strategies (fig. S9). Then, the flow cytometry analysis was performed on  $10^5$  live cells for each sample. Data were analyzed using FCS Express 6 software (De Novo Software). The percentage of mNG2-positive population of each sample was used for normalization (table S4).

### Flow cytometry analysis for the protein expression assay and antibody binding assay

Approximately  $1 \times 10^6$  cells that carried the selected SARS-CoV-2 S NTD mutant were washed with  $1 \times$  PBS, harvested with warm versene, and pelleted via centrifugation at 300g for 5 min at room temperature. The cells were resuspended in the FACS buffer. Subsequently, cells were incubated with 5  $\mu$ g/ml of the selected antibodies at 4°C with gentle shaking for 1 hour. Cells were then washed once with ice-cold FACS buffer and incubated with PE anti-human IgG Fc (2  $\mu$ g/ml; BioLegend, catalog no. 410708) at 4°C with gentle shaking in the dark for 1 hour. Cells were washed once, pelleted via centrifugation at 300g for 5 min at room temperature, and resuspended in ice-cold FACS buffer. LSRII flow cytometry (BD Biosciences) was used to measure the PE signal of each sample. Negative controls were measured first to set up proper gating strategies (figs. S8 and S10). Then, the flow cytometry analysis was performed on  $10^5$  singlets for each sample. Data were analyzed using FCS Express 6 software (De Novo Software).

### Normalization of the expression assay results

The MFI of the entire population was recorded for each sample, followed by the normalization as previously described (42). For a given sample  $i$ , the following equation was used to compute the normalized expression (NE)

$$NE = \frac{MFI_i - MFI_{\text{negative control}}}{MFI_{\text{wildtype}} - MFI_{\text{negative control}}} \quad (6)$$

Normalizations were performed for each sample within a given biological replicate (table S4).

### Recombinant expression and purification of soluble S protein

SARS-CoV-2 S ectodomain with the PRRA motif in the furin cleavage site deleted and mutations K986P/V987P, which are known to stabilize the prefusion conformation and increase expression (26, 48), was cloned into a pHCMV3 vector. The S ectodomain construct contained a trimerization domain and a 6 $\times$ His-tag at the C terminus. Expi293F cells (Gibco), which were maintained using Expi293 expression medium (Gibco), were used to express soluble S protein. Briefly, 25  $\mu$ g of the plasmid was transfected into 25 ml of Expi293F cells at  $3 \times 10^6$  cells ml<sup>-1</sup> using the ExpiFectamine 293 Transfection Kit (Thermo Fisher Scientific) following the manufacturer's instructions. Transfected cells were then incubated at 37°C, 8% CO<sub>2</sub> and shaking at 125 rpm for 6 days. Cell cultures were then harvested and centrifuged at 4000g at 4°C for 15 min. The supernatant was clarified using a 0.22- $\mu$ m polyethersulfone filter (Millipore). S protein in the clarified supernatant was then purified using Nickel Sepharose Excel resin (Cytiva), with 20 mM imidazole in PBS as wash buffer, and 300 mM imidazole in PBS as elution buffer. Three rounds of 2-ml elutions were performed. The eluted protein was then concentrated and analyzed by SDS-polyacrylamide gel electrophoresis reducing gels (Bio-Rad; fig. S6A). Concentrated protein solution was further purified using Superdex 200 XK 16/100 size exclusion column (Cytiva) in 20 mM tris-HCl (pH 8.0) and 150 mM NaCl (fig. S6B). Selected elution fractions were combined and concentrated. Final protein concentration was measured using NanoDrop One (Thermo Fisher Scientific).

### Protein thermostability assay

Five micrograms of purified protein was mixed with 5 $\times$  SYPRO orange (Thermo Fisher Scientific) in 20 mM tris-HCl (pH 8.0), 150 mM NaCl at a final volume of 25  $\mu$ l. The sample mixture was then transferred into an optically clear PCR tube (VWR). SYPRO orange fluorescence data in relative fluorescence unit (RFU) was collected from 10° to 95°C using CFX Connect Real-Time PCR Detection System (Bio-Rad). The temperature corresponding to the lowest point of the first derivative,  $-d(\text{RFU})/dT$ , was defined as the melting temperature ( $T_m$ ). Data were analyzed using OriginPro 2020b (Origin Lab). Raw data are shown in table S5.

### SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <https://science.org/doi/10.1126/sciadv.add7221>

[View/request a protocol for this paper from Bio-protocol.](#)

### REFERENCES AND NOTES

1. P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, H.-D. Chen, J. Chen, Y. Luo, H. Guo, R.-D. Jiang, M.-Q. Liu, Y. Chen, X.-R. Shen, X. Wang, X.-S. Zheng, K. Zhao, Q.-J. Chen, F. Deng, L.-L. Liu, B. Yan, F.-X. Zhan, Y.-Y. Wang, G.-F. Xiao, Z.-L. Shi, A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
2. N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G. F. Gao, W. Tan, A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* **382**, 727–733 (2020).



3. M. Letko, A. Marzi, V. Munster, Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat. Microbiol.* **5**, 562–569 (2020).
4. A. C. Walls, Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, D. Veessler, Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **181**, 281–292.e6 (2020).
5. W. R. Strohl, Z. Ku, Z. An, S. F. Carroll, B. A. Keyt, L. M. Strohl, Passive immunotherapy against SARS-CoV-2: From plasma-based therapy to single potent antibodies in the race to stay ahead of the variants. *BioDrugs* **36**, 231–323 (2022).
6. M. Yuan, H. Liu, N. C. Wu, I. A. Wilson, Recognition of the SARS-CoV-2 receptor binding domain by neutralizing antibodies. *Biochem. Biophys. Res. Commun.* **538**, 192–203 (2021).
7. D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C.-L. Hsieh, O. Abiona, B. S. Graham, J. S. McLellan, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263 (2020).
8. Y. Huang, C. Yang, X. Xu, W. Xu, S. Liu, Structural and functional properties of SARS-CoV-2 spike protein: Potential antiviral drug development for COVID-19. *Acta Pharmacol. Sin.* **41**, 1141–1149 (2020).
9. L. Premkumar, B. Segovia-Chumbez, R. Jodi, D. R. Martinez, R. Raut, A. J. Markmann, C. Cornaby, L. Bartelt, S. Weiss, Y. Park, C. E. Edwards, E. Weimer, E. M. Scherer, N. Roupheal, S. Edupuganti, D. Weiskopf, L. V. Tse, Y. J. Hou, D. Margolis, A. Sette, M. H. Collins, J. Schmitz, R. S. Baric, A. M. de Silva, The receptor-binding domain of the viral spike protein is an immunodominant and highly specific target of antibodies in SARS-CoV-2 patients. *Sci. Immunol.* **5**, eabc8413 (2020).
10. L. Piccoli, Y.-J. Park, M. A. Tortorici, N. Czudnochowski, A. C. Walls, M. Beltramello, C. Silacci-Fregni, D. Pinto, L. E. Rosen, J. E. Bowen, O. J. Acton, S. Jaconi, B. Guarino, A. Minola, F. Zatta, N. Sprugasci, J. Bassi, A. Peter, A. De Marco, J. C. Nix, F. Mele, S. Jovic, B. F. Rodriguez, S. V. Gupta, F. Jin, G. Piumatti, G. L. Presti, A. F. Pellanda, M. Gioggero, M. Tarkowski, M. S. Pizzuto, E. Cameroni, C. Havenar-Daughton, M. Smithy, D. Hong, V. Lepori, E. Albanese, A. Ceschi, E. Bernasconi, L. Elzi, P. Ferrari, C. Garzoni, A. Riva, G. Snell, F. Fallusto, K. Fink, H. W. Virgin, A. Lanzavecchia, D. Corti, D. Veessler, Mapping neutralizing and immunodominant sites on the SARS-CoV-2 spike receptor-binding domain by structure-guided high-resolution serology. *Cell* **183**, 1024–1042.e21 (2020).
11. A. J. Greaney, A. N. Loes, K. H. D. Crawford, T. N. Starr, K. D. Malone, H. Y. Chu, J. D. Bloom, Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe* **29**, 463–476.e6 (2021).
12. X. Chi, R. Yan, J. Zhang, G. Zhang, Y. Zhang, M. Hao, Z. Zhang, P. Fan, Y. Dong, Y. Yang, Z. Chen, Y. Guo, J. Zhang, Y. Li, X. Song, Y. Chen, L. Xia, L. Fu, L. Hou, J. Xu, C. Yu, J. Li, Q. Zhou, W. Chen, A neutralizing human antibody binds to the N-terminal domain of the spike protein of SARS-CoV-2. *Science* **369**, 650–655 (2020).
13. L. Liu, P. Wang, M. S. Nair, J. Yu, M. Rapp, Q. Wang, Y. Luo, J. F.-W. Chan, V. Sahi, A. Figueroa, X. V. Guo, G. Cerutti, J. Bimela, J. Gorman, T. Zhou, Z. Chen, K.-Y. Yuen, P. D. Kwong, J. G. Sodroski, M. T. Yin, Z. Sheng, Y. Huang, L. Shapiro, D. D. Ho, Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. *Nature* **584**, 450–456 (2020).
14. M. McCallum, A. De Marco, F. A. Lempp, M. A. Tortorici, D. Pinto, A. C. Walls, M. Beltramello, A. Chen, Z. Liu, F. Zatta, S. Zepeda, J. di Iulio, J. E. Bowen, M. Montiel-Ruiz, J. Zhou, L. E. Rosen, S. Bianchi, B. Guarino, C. S. Fregni, R. Abdelnabi, S.-Y. C. Foo, P. W. Rothlauf, L.-M. Bloyet, F. Benigni, E. Cameroni, J. Neyts, A. Riva, G. Snell, A. Telenti, S. P. J. Whelan, H. W. Virgin, D. Corti, M. S. Pizzuto, D. Veessler, N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell* **184**, 2332–2347.e16 (2021).
15. N. Suryadevara, S. Shrihari, P. Gilchuk, L. A. VanBlargan, E. Binshtein, S. J. Zost, R. S. Nargi, R. E. Sutton, E. S. Winkler, E. C. Chen, M. E. Fouch, E. Davidson, B. J. Doranz, R. E. Chen, P.-Y. Shi, R. H. Carnahan, L. B. Thackray, M. S. Diamond, J. E. Crowe, Neutralizing and protective human monoclonal antibodies recognizing the N-terminal domain of the SARS-CoV-2 spike protein. *Cell* **184**, 2316–2331.e15 (2021).
16. Z. Wang, F. Muecksch, A. Cho, C. Gaebler, H.-H. Hoffmann, V. Ramos, S. Zong, M. Cipolla, B. Johnson, F. Schmidt, J. DaSilva, E. Bednarski, T. Ben Tanfous, R. Raspe, K. Yao, Y. E. Lee, T. Chen, M. Turroja, K. G. Milard, J. Dizon, A. Kaczynska, A. Gazumyan, T. Y. Oliveira, C. M. Rice, M. Caskey, P. D. Bieniasz, T. Hatziioannou, C. O. Barnes, M. C. Nussenzweig, Analysis of memory B cells identifies conserved neutralizing epitopes on the N-terminal domain of variant SARS-CoV-2 spike proteins. *Immunity* **55**, 998–1012.e8 (2022).
17. G. Cerutti, Y. Guo, T. Zhou, J. Gorman, M. Lee, M. Rapp, E. R. Reddem, J. Yu, F. Bahna, J. Bimela, Y. Huang, P. S. Katsamba, L. Liu, M. S. Nair, R. Rawi, A. S. Olia, P. Wang, B. Zhang, G.-Y. Chuang, D. D. Ho, Z. Sheng, P. D. Kwong, L. Shapiro, Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. *Cell Host Microbe* **29**, 819–833.e7 (2021).
18. K. A. Wolf, J. C. Kwan, J. P. Kamil, Structural dynamics and molecular evolution of the SARS-CoV-2 spike protein. *MBio* **13**, e02030-21 (2022).
19. W.-S. Lin, I.-C. Chen, H.-C. Chen, Y.-C. Lee, S.-C. Wu, Glycan masking of epitopes in the NTD and RBD of the spike protein elicits broadly neutralizing antibodies against SARS-CoV-2 variants. *Front. Immunol.* **12**, 795741 (2021).
20. M. W. Tenforde, W. H. Self, M. Gaglani, A. A. Ginde, D. J. Douin, H. K. Talbot, J. D. Casey, N. M. Mohr, A. Zepeski, T. McNeal, S. Ghamamed, K. W. Gibbs, D. C. Files, D. N. Hager, A. Shehu, M. E. Prekker, A. E. Frosch, M. N. Gong, A. Mohamed, N. J. Johnson, V. Srinivasan, J. S. Steingrub, I. D. Peltan, S. M. Brown, E. T. Martin, A. S. Monto, A. Khan, C. L. Hough, L. W. Busse, A. Duggal, J. G. Wilson, N. Qadir, S. Y. Chang, C. Mallow, C. Rivas, H. M. Babcock, J. H. Kwon, M. C. Exline, M. Botros, A. S. Lauring, N. I. Shapiro, N. Halasa, J. D. Chappell, C. G. Grijalva, T. W. Rice, I. D. Jones, W. B. Stubblefield, A. Baughman, K. N. Womack, J. P. Rhoads, C. J. Lindsell, K. W. Hart, Y. Zhu, K. Adams, D. Surie, M. L. McMorrow, M. M. Patel, I. V. Y. Network, Effectiveness of mRNA vaccination in preventing COVID-19-associated invasive mechanical ventilation and death - United States, March 2021-January 2022. *MMWR Morb. Mortal. Wkly Rep.* **71**, 459–465 (2022).
21. S. Y. Tartof, J. M. Slezak, H. Fischer, V. Hong, B. K. Ackerson, O. N. Ranasinghe, T. B. Frankland, O. A. Ogun, J. M. Zamparo, S. Gray, S. R. Valluri, K. Pan, F. J. Angulo, L. Jodar, J. M. McLaughlin, Effectiveness of mRNA BNT162b2 COVID-19 vaccine up to 6 months in a large integrated health system in the USA: A retrospective cohort study. *Lancet* **398**, 1407–1416 (2021).
22. P. T. Heath, E. P. Galiza, D. N. Baxter, M. Boffito, D. Browne, F. Burns, D. R. Chadwick, R. Clark, C. Cosgrove, J. Galloway, A. L. Goodman, A. Heer, A. Higham, S. Iyengar, A. Jamal, C. Jeanes, P. A. Kalra, C. Kyriakidou, D. F. McAuley, A. Meyrick, A. M. Minassian, J. Minton, P. Moore, I. Munsoor, H. Nicholls, O. Osanlou, J. Packham, C. H. Pretswell, A. S. F. Ramos, D. Saralaya, R. P. Sheridan, R. Smith, R. L. Soiza, P. A. Swift, E. C. Thomson, J. Turner, M. E. Viljoen, G. Albert, I. Cho, F. Dubovsky, G. Glenn, J. Rivers, A. Robertson, K. Smith, S. Toback; 2019nCoV-302 Study Group, Safety and efficacy of NVX-CoV2373 COVID-19 vaccine. *N. Engl. J. Med.* **385**, 1172–1183 (2021).
23. S. Sridhar, A. Joaquin, M. I. Bonaparte, A. Bueso, A.-L. Chabanon, A. Chen, R. M. Chic, D. Diemert, B. J. Essink, B. Fu, N. A. Grunenber, H. Janoszyk, M. C. Keefer, D. M. R. M, Y. Meng, N. L. Michael, S. S. Munsiff, O. Ogbuagu, V. N. Raabe, R. Severance, E. Rivas, N. Romanyak, N. G. Roupheal, L. Schuerman, L. D. Sher, S. R. Walsh, J. White, D. von Barbier, G. de Bruyn, R. Canter, M.-H. Grillet, M. Keshtkar-Jahromi, M. Koutsoukos, D. Lopez, R. Masotti, S. Mendoza, C. Moreau, M. A. Cerejido, S. Ramirez, A. Said, F. Tavares-Da-Silva, J. Shi, T. Tong, J. Treanor, C. A. Diazgranados, S. Savarino, Safety and immunogenicity of an AS03-adjuvanted SARS-CoV-2 recombinant protein vaccine (CoV2 preS dTM) in healthy adults: Interim findings from a phase 2, randomised, dose-finding, multicentre study. *Lancet Infect. Dis.* **22**, 636–648 (2022).
24. J. Pollet, W.-H. Chen, U. Strych, Recombinant protein vaccines, a proven approach against coronavirus pandemics. *Adv. Drug Deliv. Rev.* **170**, 71–82 (2021).
25. T. Schlake, A. Thess, M. Fotin-Mlecsek, K.-J. Kallen, Developing mRNA-vaccine technologies. *RNA Biol.* **9**, 1319–1330 (2012).
26. J. Juraszek, L. Rutten, S. Blokland, P. Bouchier, R. Voorzaat, T. Ritschel, M. J. G. Bakkers, L. L. R. Renault, J. P. M. Langedijk, Stabilizing the closed SARS-CoV-2 spike trimer. *Nat. Commun.* **12**, 244 (2021).
27. C.-L. Hsieh, J. A. Goldsmith, J. M. Schaub, A. M. DiVenere, H.-C. Kuo, K. Javanmardi, K. C. Le, D. Wrapp, A. G. Lee, Y. Liu, C.-W. Chou, P. O. Byrne, C. K. Hjorth, N. V. Johnson, J. Ludes-Meyers, A. W. Nguyen, J. Park, N. Wang, D. Amengor, J. J. Lavinder, G. C. Ippolito, J. A. Maynard, I. J. Finkelstein, J. S. McLellan, Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science* **369**, 1501–1505 (2020).
28. E. Olmedillas, C. J. Mann, W. Peng, Y.-T. Wang, R. D. Avalos, D. Bedinger, K. Valentine, N. Shafee, S. L. Schendel, M. Yuan, G. Lang, R. Rouet, D. Christ, W. Jiang, I. A. Wilson, T. Germann, S. Shresta, J. Snijder, E. O. Saphire, Structure-based design of a highly stable, covalently-linked SARS-CoV-2 spike trimer with improved structural properties and immunogenicity. *bioRxiv* 2021.05.06.441046 [Preprint]. 6 May 2021. <https://doi.org/10.1101/2021.05.06.441046>.
29. D. Ellis, N. Brunette, K. H. D. Crawford, A. C. Walls, M. N. Pham, C. Chen, K.-L. Herpold, B. Fiala, M. Murphy, D. Pettie, J. C. Kraft, K. D. Malone, M. J. Navarro, C. Ogohara, E. Kepl, R. Ravichandran, C. Sydeman, M. Ahlrichs, M. Johnson, A. Blackstone, L. Carter, T. N. Starr, A. J. Greaney, K. K. Lee, D. Veessler, J. D. Bloom, N. P. King, Stabilization of the SARS-CoV-2 spike receptor-binding domain using deep mutational scanning and structure-based design. *Front. Immunol.* **12**, 710263 (2021).
30. D. M. Fowler, S. Fields, Deep mutational scanning: A new style of protein science. *Nat. Methods* **11**, 801–807 (2014).
31. A. J. Greaney, T. N. Starr, C. O. Barnes, Y. Weisblum, F. Schmidt, M. Caskey, C. Gaebler, A. Cho, M. Agudelo, S. Finklin, Z. Wang, D. Poston, F. Muecksch, T. Hatziioannou, P. D. Bieniasz, D. F. Robbiani, M. C. Nussenzweig, P. J. Bjorkman, J. D. Bloom, Mapping mutations to the SARS-CoV-2 RBD that escape binding by different classes of antibodies. *Nat. Commun.* **12**, 4196 (2021).
32. A. J. Greaney, T. N. Starr, J. D. Bloom, An antibody-escape estimator for mutations to the SARS-CoV-2 receptor-binding domain. *Virus Evol.* **8**, veac021 (2022).
33. T. N. Starr, N. Czudnochowski, Z. Liu, F. Zatta, Y.-J. Park, A. Addetia, D. Pinto, M. Beltramello, P. Hernandez, A. J. Greaney, R. Marzi, W. G. Glass, I. Zhang, A. S. Dingsen, J. E. Bowen, M. A. Tortorici, A. C. Walls, J. A. Wojcechowskyj, A. De Marco, L. E. Rosen, J. Zhou, M. Montiel-Ruiz, H. Kaiser, J. R. Dillen, H. Tucker, J. Bassi, C. Silacci-Fregni, M. P. Housley,



- J. di Iulio, G. Lombardo, M. Agostini, N. Sprugasci, K. Culp, S. Jaconi, M. Meury, E. Dellota Jr., R. Abdelnabi, S.-Y. C. Foo, E. Cameroni, S. Stumpf, T. I. Croll, J. C. Nix, C. Havenar-Daughton, L. Piccoli, F. Benigni, J. Neyts, A. Telenti, F. A. Lempp, M. S. Pizzuto, J. D. Chodera, C. M. Hebnar, H. W. Virgin, S. P. J. Whelan, D. Veessler, D. Corti, J. D. Bloom, G. Snell, SARS-CoV-2 RBD antibodies that maximize breadth and resistance to escape. *Nature* **597**, 97–102 (2021).
34. T. N. Starr, A. J. Greaney, S. K. Hilton, D. Ellis, K. H. D. Crawford, A. S. Dingens, M. J. Navarro, J. E. Bowen, M. A. Tortorici, A. C. Walls, N. P. King, D. Veessler, J. D. Bloom, Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* **182**, 1295–1310.e20 (2020).
35. T. N. Starr, A. J. Greaney, A. S. Dingens, J. D. Bloom, Complete map of SARS-CoV-2 RBD mutations that escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016. *Cell Rep. Med.* **2**, 100255 (2021).
36. T. N. Starr, A. J. Greaney, A. Addetia, W. W. Hannon, M. C. Choudhary, A. S. Dingens, J. Z. Li, J. D. Bloom, Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science* **371**, 850–854 (2021).
37. K. A. Matreyek, J. J. Stephany, D. M. Fowler, A platform for functional assessment of large variant libraries in mammalian cells. *Nucleic Acids Res.* **45**, e102 (2017).
38. N. Shukla, S. M. Roelle, V. G. Suzart, A. M. Bruchez, K. A. Matreyek, Mutants of human ACE2 differentially promote SARS-CoV and SARS-CoV-2 spike mediated infection. *PLOS Pathog.* **17**, e1009715 (2021).
39. P. Zhou, M. Yuan, G. Song, N. Beutler, N. Shaabani, D. Huang, W. He, X. Zhu, S. Callaghan, P. Yong, F. Anzanello, L. Peng, J. Ricketts, M. Parren, E. Garcia, S. A. Rawlings, D. M. Smith, D. Nemazee, J. R. Tejjara, T. F. Rogers, I. A. Wilson, D. R. Burton, R. Andrabai, A human antibody reveals a conserved site on beta-coronavirus spike proteins and confers protection against SARS-CoV-2 infection. *Sci. Transl. Med.* **14**, eab91215 (2022).
40. C. A. Olson, N. C. Wu, R. Sun, A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. *Curr. Biol.* **24**, 2643–2651 (2014).
41. K. A. Matreyek, L. M. Starita, J. J. Stephany, B. Martin, M. A. Chiasson, V. E. Gray, M. Kircher, A. Khechaduri, J. N. Dines, R. J. Hause, S. Bhatia, W. E. Evans, M. V. Relling, W. Yang, J. Shendure, D. M. Fowler, Multiplex assessment of protein variant abundance by massively parallel sequencing. *Nat. Genet.* **50**, 874–882 (2018).
42. K. K. Chan, T. J. C. Tan, K. K. Narayanan, E. Procko, An engineered decoy receptor for SARS-CoV-2 broadly binds protein S sequence variants. *Sci. Adv.* **7**, eabf1738 (2021).
43. M. McCallum, A. C. Walls, K. R. Sprouse, J. E. Bowen, L. E. Rosen, H. V. Dang, A. De Marco, N. Franko, S. W. Tilles, J. Logue, M. C. Miranda, M. Ahlrichs, L. Carter, G. Snell, M. S. Pizzuto, H. Y. Chu, W. C. Van Voorhis, D. Corti, D. Veessler, Molecular basis of immune evasion by the Delta and Kappa SARS-CoV-2 variants. *Science* **374**, 1621–1626 (2021).
44. Y. Shu, J. McCauley, GISAD: Global initiative on sharing all influenza data—From vision to reality. *Euro Surveill.* **22**, 30494 (2017).
45. A. G. Wrobel, D. J. Benton, P. Xu, C. Roustan, S. R. Martin, P. B. Rosenthal, J. J. Skehel, S. J. Gamblin, SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. *Nat. Struct. Mol. Biol.* **27**, 763–767 (2020).
46. A. Rosa, V. E. Pye, C. Graham, L. Muir, J. Seow, K. W. Ng, N. J. Cook, C. Rees-Spear, E. Parker, M. S. D. Santos, C. Rosadas, A. Susana, H. Rhys, A. Nans, L. Masino, C. Roustan, E. Christodoulou, R. Ulferts, A. G. Wrobel, C.-E. Short, M. Fertleman, R. W. Sanders, J. Heaney, M. Spyer, S. Kjær, A. Riddell, M. H. Malim, R. Beale, J. I. MacRae, G. P. Taylor, E. Nastouli, M. J. van Gils, P. B. Rosenthal, M. Pizzato, M. O. McClure, R. S. Tedder, G. Kassiotis, L. E. McCoy, K. J. Doores, P. Cherepanov, SARS-CoV-2 can recruit a heme metabolite to evade antibody immunity. *Sci. Adv.* **7**, eabg7607 (2021).
47. E. H. Kellogg, A. Leaver-Fay, D. Baker, Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins* **79**, 830–838 (2011).
48. R. W. Sanders, J. P. Moore, Virus vaccines: Proteins prefer prolines. *Cell Host Microbe* **29**, 327–333 (2021).
49. S. Feng, S. Sekine, V. Pessino, H. Li, M. D. Leonetti, B. Huang, Improved split fluorescent proteins for endogenous protein labeling. *Nat. Commun.* **8**, 370 (2017).
50. M. Yuan, D. Huang, C.-C. D. Lee, N. C. Wu, A. M. Jackson, X. Zhu, H. Liu, L. Peng, M. J. van Gils, R. W. Sanders, D. R. Burton, S. M. Reincke, H. Prüss, J. Kreye, D. Nemazee, A. B. Ward, I. A. Wilson, Structural and functional ramifications of antigenic drift in recent SARS-CoV-2 variants. *Science* **373**, 818–823 (2021).
51. M. Claireaux, T. G. Daniels, M. de Gast, J. Han, D. Guerra, G. Kerster, B. D. van Schaik, A. Jongejan, A. I. Schriek, M. Grobbee, P. J. Brouwer, K. van der Straten, Y. Aldon, J. Capella-Pujol, J. L. Snitselaar, W. Olijhoek, A. Aartse, M. Brinkkemper, I. Bontjer, J. A. Burger, M. Poniman, T. P. Bijl, J. L. Torres, J. Cops, I. C. Martin, S. W. de Taeye, G. J. de Bree, A. B. Ward, K. Slieden, A. H. van Kampen, P. D. Moerland, R. W. Sanders, M. J. van Gils, A public antibody class recognizes an S2 epitope exposed on open conformations of SARS-CoV-2 spike. *Nat. Commun.* **13**, 4539 (2022).
52. J. W. Yewdell, Antigenic drift: Understanding COVID-19. *Immunity* **54**, 2681–2687 (2021).
53. W. T. Harvey, A. M. Carabelli, B. Jackson, R. K. Gupta, E. C. Thomson, E. M. Harrison, C. Ludden, R. Reeve, A. Rambaut, S. J. Peacock, D. L. Robertson, SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* **19**, 409–424 (2021).
54. A. Telenti, A. Arvin, L. Corey, D. Corti, M. S. Diamond, A. Garcia-Sastre, R. F. Garry, E. C. Holmes, P. S. Pang, H. W. Virgin, After the pandemic: Perspectives on the future trajectory of COVID-19. *Nature* **596**, 495–504 (2021).
55. P. Wang, M. S. Nair, L. Liu, S. Iketani, Y. Luo, Y. Guo, M. Wang, J. Yu, B. Zhang, P. D. Kwong, B. S. Graham, J. R. Mascola, J. Y. Chang, M. T. Yin, M. Sobieszczyk, C. A. Kyratsous, L. Shapiro, Z. Sheng, Y. Huang, D. D. Ho, Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* **593**, 130–135 (2021).
56. R. E. Chen, X. Zhang, J. B. Case, E. S. Winkler, Y. Liu, L. A. VanBlargan, J. Liu, J. M. Errico, X. Xie, N. Suryadevara, P. Gilchuk, S. J. Zost, S. Tahan, L. Droit, J. S. Turner, W. Kim, A. J. Schmitz, M. Thapa, D. Wang, A. C. M. Boon, R. M. Presti, J. A. O'Halloran, A. H. J. Kim, P. Deepak, D. Pinto, D. H. Fremont, J. E. Crowe, D. Corti, H. W. Virgin, A. H. Ellebedy, P.-Y. Shi, M. S. Diamond, Resistance of SARS-CoV-2 variants to neutralization by monoclonal and serum-derived polyclonal antibodies. *Nat. Med.* **27**, 717–726 (2021).
57. M. McCallum, J. Bassi, A. De Marco, A. Chen, A. C. Walls, J. Di Iulio, M. A. Tortorici, M.-J. Navarro, C. Silacci-Fregni, C. Saliba, K. R. Sprouse, M. Agostini, D. Pinto, K. Culp, S. Bianchi, S. Jaconi, E. Cameroni, J. E. Bowen, S. W. Tilles, M. S. Pizzuto, S. B. Gustalla, G. Bona, A. F. Pellanda, C. Garzoni, W. C. Van Voorhis, L. E. Rosen, G. Snell, A. Telenti, H. W. Virgin, L. Piccoli, D. Corti, D. Veessler, SARS-CoV-2 immune evasion by the B.1.427/B.1.429 variant of concern. *Science* **373**, 648–654 (2021).
58. P. Mlcochova, S. A. Kemp, M. S. Dhar, G. Papa, B. Meng, I. A. T. M. Ferreira, R. Datir, D. A. Collier, A. Albecka, S. Singh, R. Pandey, J. Brown, J. Zhou, N. Goonawardane, S. Mishra, C. Whittaker, T. Mellan, R. Marwal, M. Datta, S. Sengupta, K. Ponnusamy, V. S. Radhakrishnan, A. Abdullahi, O. Charles, P. Chattopadhyay, P. Devi, D. Caputo, T. Peacock, C. Watal, N. Goel, A. Satwik, R. Vaishya, M. Agarwal, Indian SARS-CoV-2 Genomics Consortium (INSACOG); Genotype to Phenotype Japan (G2P-Japan) Consortium; CITIID-NIHR BioResource COVID-19 Collaboration, A. Mavousian, J. H. Lee, J. Bassi, C. Silacci-Fegni, C. Saliba, D. Pinto, T. Irie, I. Yoshida, W. L. Hamilton, K. Sato, S. Bhatt, S. Flaxman, L. C. James, D. Corti, L. Piccoli, W. S. Barclay, P. Rakshit, A. Agrawal, R. K. Gupta, SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature* **599**, 114–119 (2021).
59. E. Cameroni, J. E. Bowen, L. E. Rosen, C. Saliba, S. K. Zepeda, K. Culp, D. Pinto, L. A. VanBlargan, A. De Marco, J. di Iulio, F. Zatta, H. Kaiser, J. Noack, N. Farhat, N. Czudnochowski, C. Havenar-Daughton, K. R. Sprouse, J. R. Dillen, A. E. Powell, A. Chen, C. Maher, L. Yin, D. Sun, L. Soriaga, J. Bassi, C. Silacci-Fregni, C. Gustafsson, N. M. Franko, J. Logue, N. T. Iqbal, I. Mazzitelli, J. Geffner, R. Grifantini, H. Chu, A. Gori, A. Riva, O. Giannini, A. Ceschi, P. Ferrari, P. E. Cippà, A. Franzetti-Pellanda, C. Garzoni, P. J. Halfmann, Y. Kawoaka, C. Hebnar, L. A. Purcell, L. Piccoli, M. S. Pizzuto, A. C. Walls, M. S. Diamond, A. Telenti, H. W. Virgin, A. Lanzavecchia, G. Snell, D. Veessler, D. Corti, Broadly neutralizing antibodies overcome SARS-CoV-2 Omicron antigenic shift. *Nature* **602**, 664–670 (2022).
60. Y. Singh, N. K. Fuloria, S. Fuloria, V. Subramaniyan, D. U. Meenakshi, S. Chakravarthi, U. Kumari, N. Joshi, G. Gupta, N-terminal domain of SARS CoV-2 spike protein mutation associated reduction in effectiveness of neutralizing antibody with vaccinated individuals. *J. Med. Virol.* **93**, 5726–5728 (2021).
61. S. M.-C. Gobeil, K. Janowska, S. McDowell, K. Mansouri, R. Parks, V. Stalls, M. F. Kopp, K. Manne, D. Li, K. Wiehe, K. O. Saunders, R. J. Edwards, B. Korber, B. F. Haynes, R. Henderson, P. Acharya, Effect of natural mutations of SARS-CoV-2 on spike structure, conformation, and antigenicity. *Science* **373**, eabi6226 (2021).
62. S.-M. Lok, An NTD supersite of attack. *Cell Host Microbe* **29**, 744–746 (2021).
63. K. Javanmardi, T. H. Segall-Shapiro, C.-W. Chou, D. R. Boutz, R. J. Olsen, X. Xie, H. Xia, P.-Y. Shi, C. D. Johnson, A. Annappareddy, S. Weaver, J. M. Musser, A. D. Ellington, I. J. Finkelstein, J. D. Gollihar, Antibody escape and cryptic cross-domain stabilization in the SARS-CoV-2 Omicron spike protein. *Cell Host Microbe* **30**, 1242–1254.e6 (2022). <https://doi.org/10.1101/2022.04.18.488614>.
64. S. Wang, Z. Qiu, Y. Hou, X. Deng, W. Xu, T. Zhou, P. Wu, S. Xie, W. Bian, C. Zhang, Z. Sun, K. Liu, C. Han, A. Lin, S. Jiang, Y. Xie, Q. Zheng, L. Lu, J. Huang, X. Li, AXL is a candidate receptor for SARS-CoV-2 that promotes infection of pulmonary and bronchial epithelial cells. *Cell Res.* **31**, 126–140 (2021).
65. Y. Gu, J. Cao, X. Zhang, H. Gao, Y. Wang, J. Wang, J. He, X. Jiang, J. Zhang, G. Shen, J. Yang, X. Zheng, G. Hu, Y. Zhu, S. Du, Y. Zhu, R. Zhang, J. Xu, F. Lan, D. Qu, G. Xu, Y. Zhao, D. Gao, Y. Xie, M. Luo, Z. Lu, Receptome profiling identifies KREMEN1 and ASGR1 as alternative functional receptors of SARS-CoV-2. *Cell Res.* **32**, 24–37 (2022).
66. W. T. Soh, Y. Liu, E. E. Nakayama, C. Ono, S. Torii, H. Nakagami, Y. Matsuura, T. Shioda, H. Arase, The N-terminal domain of spike glycoprotein mediates SARS-CoV-2 infection by associating with L-SIGN and DC-SIGN. *bioRxiv* 2020.11.05.369264 [Preprint]. 5 November 2020. <https://doi.org/10.1101/2020.11.05.369264>.
67. E. Qing, P. Li, L. Cooper, S. Schulz, H.-M. Jäck, L. Rong, S. Perlman, T. Gallagher, Inter-domain communication in SARS-CoV-2 spike proteins controls protease-triggered cell entry. *Cell Rep.* **39**, 110786 (2022).

68. E. Qing, T. Kicmal, B. Kumar, G. M. Hawkins, E. Timm, S. Perlman, T. Gallagher, Dynamics of SARS-CoV-2 spike proteins in cell entry: Control elements in the amino-terminal domains. *mBio* **12**, e0159021 (2021).
69. Y. Wang, R. Lei, A. Nourmohammad, N. C. Wu, Antigenic evolution of human influenza H3N2 neuraminidase is constrained by charge balancing. *eLife* **10**, e72516 (2021).
70. J. Zhang, K. Kobert, T. Flouri, A. Stamatakis, PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).
71. P. J. A. Cock, T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski, M. J. L. de Hoon, Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
72. W. Kabsch, C. Sander, Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).
73. R. P. Joosten, T. A. H. te Beek, E. Krieger, M. L. Hekkelman, R. W. W. Hoof, R. Schneider, C. Sander, G. Vriend, A series of PDB related databases for everyday needs. *Nucleic Acids Res.* **39**, D411–D419 (2011).
74. M. Z. Tien, A. G. Meyer, D. K. Sydykova, S. J. Spielman, C. O. Wilke, Maximum allowed solvent accessibilities of residues in proteins. *PLOS ONE* **8**, e80635 (2013).
75. N. C. Wu, M. Yuan, S. Bangaru, D. Huang, X. Zhu, C.-C. D. Lee, H. L. Turner, L. Peng, L. Yang, D. R. Burton, D. Nemazee, A. B. Ward, I. A. Wilson, A natural mutation between SARS-CoV-2 and SARS-CoV determines neutralization by a cross-reactive antibody. *PLOS Pathog.* **16**, e1009089 (2020).
76. S. Temmam, K. Vongphayloth, E. Baquero, S. Munier, M. Bonomi, B. Regnault, B. Douangboubpha, Y. Karami, D. Chrétien, D. Sanamxay, V. Xayaphet, P. Paphaphanh, V. Lacoste, S. Somlor, K. Lakeomany, N. Phommavanh, P. Pérot, O. Dehan, F. Amara, F. Donati, T. Bigot, M. Nilges, F. A. Rey, S. van der Werf, P. T. Brey, M. Eloit, Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. *Nature* **604**, 330–336 (2022).
77. H. Zhou, X. Chen, T. Hu, J. Li, H. Song, Y. Liu, P. Wang, D. Liu, J. Yang, E. C. Holmes, A. C. Hughes, Y. Bi, W. Shi, A novel bat coronavirus closely related to sars-cov-2 contains natural insertions at the s1/s2 cleavage site of the spike protein. *Curr. Biol.* **30**, 2196–2203.e3 (2020).
78. T. T.-Y. Lam, N. Jia, Y.-W. Zhang, M. H.-H. Shum, J.-F. Jiang, H.-C. Zhu, Y.-G. Tong, Y.-X. Shi, X.-B. Ni, Y.-S. Liao, W.-J. Li, B.-G. Jiang, W. Wei, T.-T. Yuan, K. Zheng, X.-M. Cui, J. Li, G.-Q. Pei, X. Qiang, W. Y.-M. Cheung, L.-F. Li, F.-F. Sun, S. Qin, J.-C. Huang, G. M. Leung, E. C. Holmes, Y.-L. Hu, Y. Guan, W.-C. Cao, Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* **583**, 282–285 (2020).
79. Y. Tao, S. Tong, Complete genome sequence of a severe acute respiratory syndrome-related coronavirus from Kenyan bats. *Microbiol. Resour. Announc.* **8**, e00548-19 (2019).
80. D. Ream, A. J. Kiss, NCBI/GenBank BLAST output XML parser tool (2013); [www.semanticscholar.org/paper/NCBI-%2F-GenBank-BLAST-Output-XML-Parser-Tool-Ream-Kiss/3ead0ae31b91d3096369de11f3488024f752bdc5](http://www.semanticscholar.org/paper/NCBI-%2F-GenBank-BLAST-Output-XML-Parser-Tool-Ream-Kiss/3ead0ae31b91d3096369de11f3488024f752bdc5).
81. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
82. C. Hulo, E. de Castro, P. Masson, L. Bougueleret, A. Bairoch, I. Xenarios, P. Le Mercier, ViralZone: A knowledge resource to understand virus diversity. *Nucleic Acids Res.* **39**, D576–D582 (2011).
83. J. P. G. L. M. Rodrigues, J. M. C. Teixeira, M. Trellet, A. M. J. J. Bonvin, pdb-tools: A swiss army knife for molecular structures. *F1000Research* **7**, 1961 (2018).

**Acknowledgments:** We thank M. Yuan, H. Lv, and Q. W. Teo for helpful discussion and the Roy J. Carver Biotechnology Center at the University of Illinois at Urbana-Champaign for assistance with fluorescence-activated cell sorting and next-generation sequencing.

**Funding:** This work was supported by NIH grants R00 AI139445 (to N.C.W.), DP2 AT011966 (to N.C.W.), and R01 AI167910 (to N.C.W.) and The Michelson Prizes for Human Immunology and Vaccine Research (to N.C.W.). **Author contributions:** Conceptualization: W.O.O., T.J.C.T., and N.C.W. Methodology: W.O.O., T.J.C.T., K.A.M., and N.C.W. Investigation: W.O.O., T.J.C.T., R.L., and N.C.W. Resources: G.S., R.A., and K.A.M. Formal analysis: W.O.O., T.J.C.T., C.K., and N.C.W. Visualization: W.O.O., T.J.C.T., and N.C.W. Supervision: N.C.W. Writing—original draft: W.O.O., T.J.C.T., and N.C.W. Writing—review and editing: W.O.O., T.J.C.T., R.L., G.S., C.K., R.A., K.A.M., and N.C.W. **Competing interests:** N.C.W. consults for HeliXon. The authors declare no other competing interests. **Data and materials availability:** Raw sequencing data have been submitted to the NIH Short Read Archive under accession number: BioProject PRJNA792013. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Source data and custom python/R scripts for data analysis and plotting in this study have been deposited to Zenodo (<https://doi.org/10.5281/zenodo.7066099>) and GitHub ([https://github.com/nicwulab/SARS-CoV-2\\_NTD\\_DMS](https://github.com/nicwulab/SARS-CoV-2_NTD_DMS)).

Submitted 29 June 2022

Accepted 3 October 2022

Published 23 November 2022

10.1126/sciadv.add7221